The EMBO Journal vol.14 no.11 pp.2661–2669, 1995

McrB: a prokaryotic protein specifically recognizing DNA containing modified cytosine residues

T.Krüger, C.Wild and M.Noyer-Weidner[1,2]

Max-Planck-Institut für Molekulare Genetik, Ihnestraße 73, D-14195 Berlin, Germany
[1]Present address: Walter de Gruyter & Co., Scientific Publishers, Genthiner Straße 13, D-10728 Berlin, Germany

[2]Corresponding author

Communicated by W.Schaffner

Restriction of DNA by the *Escherichia coli* K-12 McrBC restriction endonuclease, which consists of the two subunits McrB and McrC, depends on the presence of modified cytosine residues in a special constellation. From previous work by others it was known that restriction of 5-methylcytosine-containing DNA requires two methylated 5′-PuC sites separated by ~40–80 non-defined base pairs. Here we show that binding of the McrBC nuclease is mediated exclusively by the McrB subunit. McrB has a low affinity for non-methylated DNA, with which it forms low molecular weight complexes. The affinity for DNA is significantly increased, with variations depending on the sequence context, by hemi- or fully methylated 5′-PuC sites. Binding to such substrates yields high molecular weight complexes, presumably involving several McrB molecules. Methylation at unique 5′-PuC sites can be sufficient to stimulate DNA binding by McrB. As such substrates are not cleaved by the nuclease, restriction apparently requires the coordinated interaction of molecules bound to neighbouring 5′-Pu^{m}C sites. The binding properties of McrB exhibit some similarities to recently identified eukaryotic proteins interacting in a non-sequence-specific manner with DNA containing methylated 5′-CpG sequences and might point to a common molecular origin of these proteins. In addition to DNA, McrB also binds GTP, an essential cofactor in DNA restriction by McrBC. McrC neither binds to DNA nor modulates the DNA binding potential of McrB. As McrC is essential for restriction it appears to predominantly function in catalysis.

Key words: DNA methylation/DNA restriction/McrBC restriction endonuclease/5-methylcytosine/protein–DNA interaction

Introduction

Escherichia coli K-12 codes for several activities which restrict DNA dependent on the presence of modified bases in specific sequence contexts (reviewed by Noyer-Weidner and Trautner, 1993). Among these, the McrBC activity, which is potentially responsive to three different modification types, 5-methylcytosine (5mC), 5-hydroxymethylcytosine and *N*4-methylcytosine, has been investigated most intensively, both at the genetic and protein level (for a review see Raleigh, 1992).

The nuclease is encoded by two slightly overlapping genes, which constitute the *mcrBC* locus located at 99 min on the *E.coli* K-12 chromosome (Raleigh *et al.*, 1989; Ross *et al.*, 1989; Dila *et al.*, 1990). *In vivo* and *in vitro* expression of the genes leads, due to alternative translational start sites, to the synthesis of five gene products (Krüger *et al.*, 1992). Restriction depends on the combined activity of the largest gene products, a 53 kDa protein specified by *mcrB* and a 39 kDa protein encoded by *mcrC* (Dila *et al.*, 1990). The smaller peptides are not required for the restriction reaction (Sutherland *et al.*, 1992). Their function is as yet unknown.

The biochemical characterization of the cleavage of 5mC-containing DNA by the McrBC nuclease (Sutherland *et al.*, 1992) revealed an unprecedented target specificity. DNA cleavage requires two methylated 5′-PuC sites separated by ~40–80 non-defined base pairs. Hemimethylation at the defined half sites (obligatory in the case of 5′-AC sites), even when affecting opposite DNA strands, is sufficient to elicit restriction. DNA cleavage itself occurs at multiple positions between the two defined half sites. Another characteristic feature of McrBC restriction is its absolute dependence on GTP, which is apparently hydrolysed in the reaction. According to Sutherland *et al.* (1992), ATP, the cofactor of other nucleotide-dependent nucleases, exerts an inhibitory effect on McrBC restriction.

While so far no experimental evidence has been provided for the particular contribution of the individual McrB and McrC subunits to the overall functional capacity of this enzyme, some inferences based on the deduced amino acid sequences of these proteins have been made (Ross *et al.*, 1989; Dila *et al.*, 1990). DNA binding has been attributed to the McrC subunit due to its basic character (Dila *et al.*, 1990). Within this subunit sequence elements with similarity to a leucine zipper and an adenine nucleotide binding motif (Ross *et al.*, 1989) were also identified, whose functional significance remains open. The McrB subunit has been suggested to be responsible for GTP binding, as its sequence contains a tripartite motif with similarities to a motif found in many guanine nucleotide binding proteins (Dever *et al.*, 1987; Dila *et al.*, 1990).

In this communication we are concerned with assigning the individual contributions of the McrB and McrC subunits to the DNA degradation process. In particular we address questions related to the mode of DNA binding. Will the whole complex be required for binding or will binding be mediated by only one of its two subunits? Will binding, as restriction, require two appropriately spaced 5′-Pu^{m}C sites or will a unique site be sufficient? Will the efficiency and quality of binding be modulated by different numbers of 5′-Pu^{m}Cs in a given DNA fragment? We also

investigated the effect of GTP and ATP on restriction and binding, with the aim of contributing to an experimental definition of the McrBC subunits interacting with these nucleotides.

Results

Overproduction and purification of McrB/C (fusion) proteins

Initial attempts to achieve an inducible, high level overproduction of the McrB and McrC proteins, by bringing the corresponding genes under the control of the *tac* or T7 promoters, failed (data not shown). To overcome these problems, which apparently derived from low efficiency of the native translational signal sequences of the *mcrB/mcrC* genes, and to facilitate protein purification we fused the genes separately to the 3′-end of the glutathione S-transferase (GST) gene using the vectors pGEX-2T or pGEX-3X (Smith and Johnson, 1988). In the resulting constructs translation should be governed by the efficient translational signals of *gst*. Purification would use affinity chromatography of the fusion proteins with immobilized glutathione, from which the McrB/C proteins would be released by proteolytic treatment with either thrombin (pGEX-2T) or factor Xa (pGEX-3X) (Smith and Johnson, 1988).

Cloning of the *mcrB* and *mcrC* genes into vector pGEX-2T yielded plasmids pBN210 and pBN212 (for details see Materials and methods). Induction of gene expression by IPTG treatment of plasmid-carrying cells surprisingly yielded overproduced proteins with sizes of the native McrB (~53 kDa) and McrC (~39 kDa) proteins (Dila *et al.*, 1990; Krüger *et al.*, 1992), rather than the anticipated GST–McrB (predicted M_r ~79 kDa) and GST–McrC (predicted M_r ~65 kDa) fusion proteins. In addition, peptides slightly larger than the free GST carrier were detected, from which free GST could be released by thrombin cleavage (data not shown).

Sequence analysis of the *gst–mcrB/C* junctions provided an explanation for this unexpected finding: due to one or two nucleotide deletions at the fusion points, slightly extended carrier peptides were synthesized whose translation terminated in the immediate vicinity of the native translational start codons (Dila *et al.*, 1990) of the *mcrB* and *mcrC* genes respectively. Due to ribosomal mRNA scanning (Adhin and van Duin, 1990), this constellation allows translational coupling (Schümperli *et al.*, 1982) between the enlarged *gst* gene and the native *mcrB* or *mcrC* genes and, consequently, overproduction of the native McrB and McrC proteins. This result implies low efficiency of the translation signal sequences of *mcrB* and *mcrC*, consistent with the observation of Sutherland *et al.* (1992) that difficulties in the expression of the *mcrB* and *mcrC* genes could be overcome by altering their translational signal sequences.

Because of their advantageous properties in purification, we made a second attempt to generate GST–McrB/C fusions. The frameshift between *mcrB* and *gst* in pBN210 was corrected by oligonucleotide-directed mutagenesis to yield pBN211 (see Materials and methods). As McrC proved to be sensitive to thrombin cleavage in the course of the experiments described above, presumably due to secondary thrombin cleavage sites (Chang, 1985), the

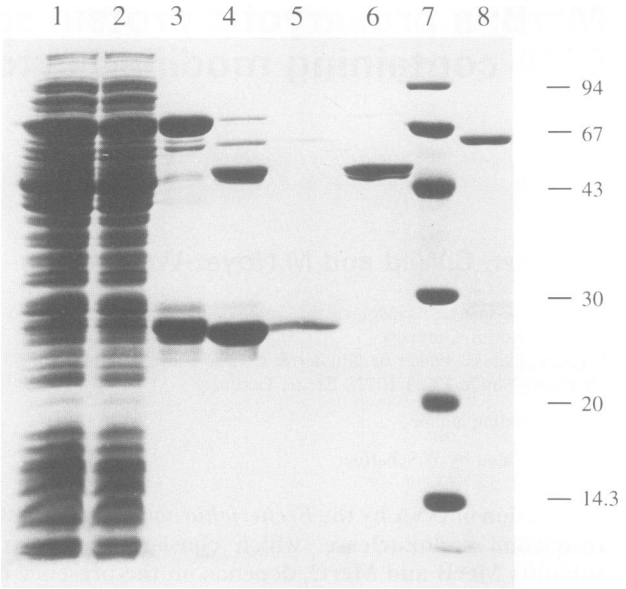

Fig. 1. Purification of the McrB protein and the final fraction of the GST–McrC protein. Proteins were separated on a 15% polyacrylamide gel and visualized with Coomassie Blue. Lanes 1–6 contain fractions of the McrB purification procedure: lane 1, supernatant of French press lysate; the lysate was passed over a glutathione–Sepharose column, the flow through of which is shown in lane 2; lane 3 contains the pool of GST–McrB-containing fractions of the glutathione–Sepharose column eluate (the band with the lowest mobility represents GST–McrB); the GST–McrB pool was subjected to thrombin cleavage (lane 4, the ~50 kDa bands represent McrB and the ~27 kDa band represents the GST carrier), followed by application onto a heparin–Sepharose column; lane 5 shows the heparin–Sepharose flow through and lane 6 the pooled and concentrated heparin–Sepharose eluate. Lane 7 contains marker proteins with sizes indicated beside the gel. Lane 8 shows the final GST–McrC fusion protein preparation after concentration. Details of the purification procedures are described in Materials and methods.

mcrC gene was cloned into vector pGEX-3X. This generated a GST–McrC fusion with a factor Xa cleavage site (plasmid pBN213). The GST–McrB and the GST–McrC fusion proteins showed the expected apparent M_r values and were visible as prominent bands within the protein profile of cell lysates after IPTG induction (Figure 1 and not shown).

Purification of the McrB protein involved affinity chromatography of the respective fusion protein on glutathione–Sepharose, followed by removal of the GST carrier by proteolysis with thrombin. Contaminating proteins and the released GST were removed by chromatography on heparin–Sepharose, to which the McrB protein bound. From 1 l of cell culture, 4 mg of fusion protein were isolated, from which ~300 μg of McrB protein could be recovered.

Figure 1 shows different purification steps. Incubation of the GST–McrB fusion protein with thrombin repeatedly yielded two peptides in the region of 50 kDa, in addition to the released 27 kDa GST protein (Figure 1, lanes 4 and 6). Determination of the N-termini of the two proteins showed that the larger product starts, as expected, with the glycine located C-terminal to the predicted thrombin cleavage site. The N-terminus of the smaller cleavage product is truncated by nine amino acids. The second cleavage point in the GST–McrB fusion protein has the

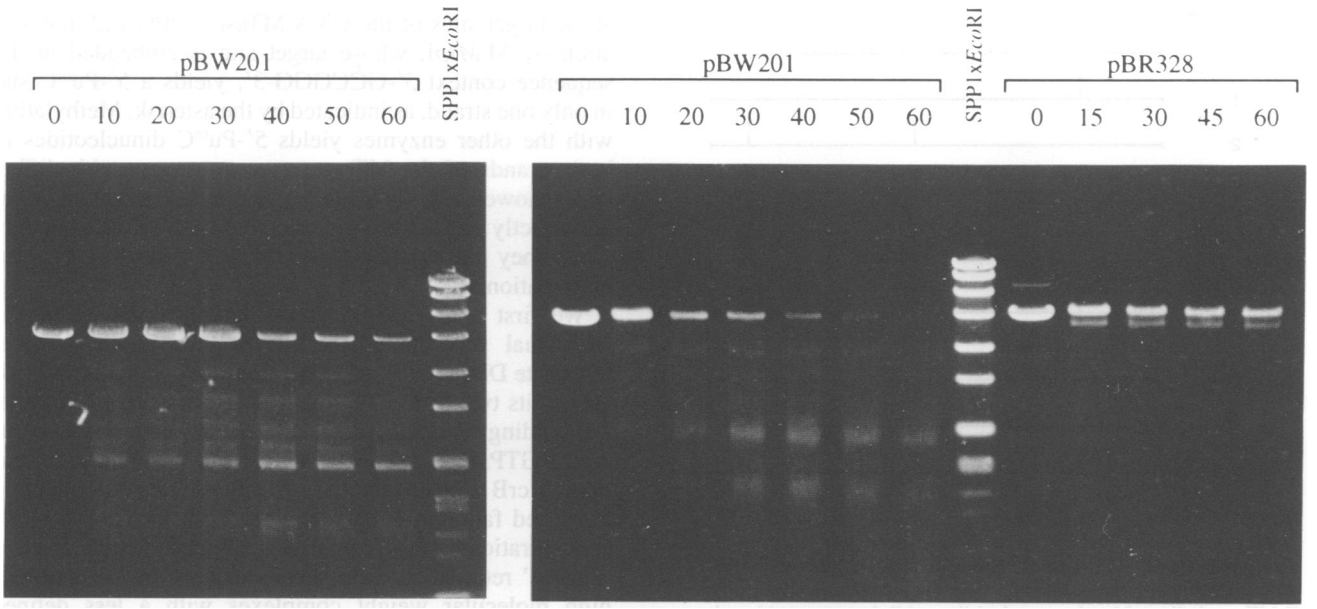

Fig. 2. Comparison of the restriction specificity of purified McrB/GST–McrC proteins and the native proteins. M.*Msp*I-specifically methylated, *Eco*RI-linearized substrate DNA (pBW201; Walter, 1990) was incubated in the presence of GTP with either a mixture of crude extracts prepared from cells harbouring pBN210 or pBN212 or with the purified McrB/GST–McrC proteins for the time periods indicated (min) (for details of the reaction conditions see Materials and methods). Incubation of the crude extracts with non-modified, *Eco*RI-linearized pBR328 DNA served as a control. The lanes SPP1×*Eco*RI contain the digestion products of DNA from phage SPP1 with R.*Eco*RI as a size marker.

sequence L-A-M-R ↓ K-A-L. This resembles a group of potentially cleavable sites, characterized by several hydrophobic amino acids immediately preceding an arginine residue, C-terminal of which cleavage occurs (Chang, 1985).

The GST–McrC fusion protein was also purified by affinity chromatography on glutathione–Sepharose. After this purification step the fusion protein was ~80–85% pure (Figure 1, lane 8). However, all attempts to release McrC in intact form from the GST–McrC fusion protein by proteolysis with factor Xa failed. Under a variety of conditions, the predominant cleavage products were peptides in the 30 kDa region, whose appearance was accompanied by a loss of McrC-specific activity (data not shown). Since McrC lacks recognition sequences for factor Xa, whose recognition stringency is much higher than that of thrombin (Nagai and Thøgersen, 1984), the reasons for its degradation by factor Xa remain unclear.

Since, for our purposes, the purity of the uncleaved GST–McrC fusion protein was sufficient, we assessed the biological activity of the GST–McrC protein and used the uncleaved fusion protein for all experiments described in the following sections.

Qualitative analysis of DNA restriction by native McrB/C proteins and GST–McrB/C fusions

We first analysed whether the native McrB and McrC proteins overproduced from pBN210 and pBN212 were functionally active. This was verified by assaying the capacity of crude extracts of Δ(*mrr-mcrBC-hsdRMS*) *E.coli* cells containing these plasmids to specifically mediate restriction of methylated DNA. Whereas separate extracts from pBN210- or pBN212-carrying cells did not cleave the DNA substrates, a combination of both extracts yielded significant methylation-dependent DNA restriction, provided the reaction buffer contained GTP. ATP neither supported nor inhibited the restriction reaction (data not shown).

Both the GST–McrB fusion protein and purified McrB obtained after thrombin cleavage of the fusion protein mediated *in vitro* restriction of modified DNA when complemented with a crude extract containing overproduced native McrC. The same observation was made when the GST–McrC fusion protein was complemented with an extract containing native McrB (data not shown).

For our analyses it was of importance to ensure that the purified fusion proteins and the McrB preparation, which consisted of two proteins with different N-termini, were equivalent in their specificity to the native proteins. We therefore compared the restriction patterns generated by incubation of methylated plasmid DNA with the purified proteins and the overproduced native McrB/C proteins contained in cell extracts. The results of restriction analyses performed with the various proteins and M.*Msp*I (^{m}CCGG)-specifically modified DNA are shown in Figure 2. A comparison of the reactions carried out with native or purified proteins revealed not only identity in the restriction patterns, but also in the intensity distribution of the individual restriction fragments. The same conclusions could be drawn from experiments involving M.*Hae*III (GG^{m}CC)-modified DNA (data not shown).

These observations indicate that the enzymatic activities of the purified McrB and GST–McrC proteins are qualitatively equivalent to those of the native McrB/C proteins.

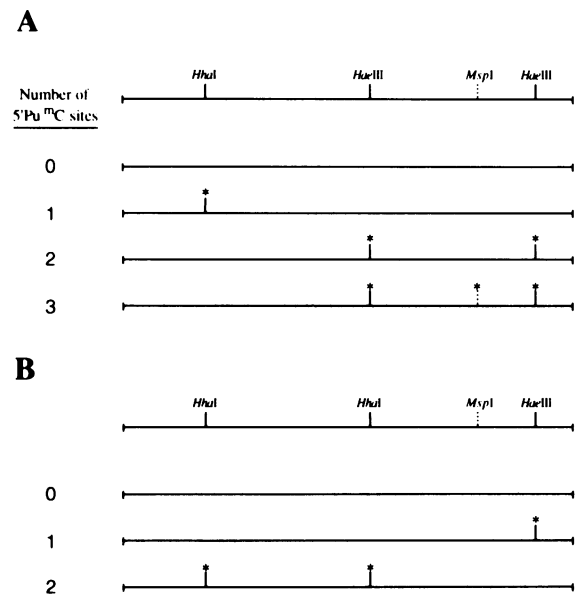

Fig. 3. Schematic representation of the DNA fragments used in DNA binding studies. (**A**) The 241 bp *Ava*II–*Acy*I restriction fragment of the vector pMS119EH (Fürste *et al.*, 1986), comprising nucleotides 162–403, is shown with the positions of relevant recognition sites for type II MTases indicated by short vertical lines. All these recognition sites contain 5′-Pu^{m}C dinucleotides. The dashed line at the M.*Msp*I site indicates the presence of a 5′-Pu^{m}C sequence in only one DNA strand. The distances between the *Hae*III sites is 89 bp, between the left-most *Hae*III site and the *Hha*I site 86 bp. Below the map of the fragment are shown differentially methylated derivatives which were used in the gel retardation experiments. The presence of 5′-Pu^{m}C is indicated by a short vertical line topped by an asterisk. (**B**) The 241 bp fragment after conversion by oligonucleotide-directed mutagenesis of the central *Hae*III site into a *Hha*I recognition sequence. All other sequences remain unaltered. Differentially methylated derivatives are shown below the map of the fragment.

It can therefore be concluded that: (i) at least one of the McrB products generated by proteolysis of the GST–McrB fusion is enzymatically active with a specificity identical to that of the native protein; (ii) the GST moiety of the GST–McrC fusion protein does not interfere with the function of the McrC subunit in qualitative respects of the restriction reaction.

While Sutherland *et al.* (1992) reported a capacity of ATP to inhibit DNA restriction by the purified McrBC proteins, no indication of an inhibitory effect of ATP was obtained in the experiments with crude extracts described above. To further address this point we incubated modified DNA with the purified McrB/GST–McrC proteins at ATP:GTP molar ratios ranging from 1:1 to 5:1. Again, no inhibition of the cleavage reaction by ATP could be observed. The same held for basically identical experiments in which the purified McrB protein was complemented with a crude extract containing native McrC (data not shown).

Characterization of the DNA binding properties of the purified McrB/C (fusion) proteins

DNA binding studies with the purified proteins were based on a gel retardation assay. A 241 bp DNA fragment, which carries several recognition sites for type II DNA methyltransferases (MTases), served as substrate (Figure 3). By separate or combined modification of this fragment with M.*Hae*III (GG^{m}CC), M.*Hha*I (G^{m}CGC) and M.*Msp*I (^{m}CCGG), different numbers of 5′-Pu^{m}C dinucleotides could be generated at defined positions. Note also subtle differences in the constellation of 5′-Pu^{m}C sites generated at the target sites of the DNA MTases employed: methylation by M.*Msp*I, whose target site is embedded in the sequence context 5′-GCCGGG-3′, yields a 5′-Pu^{m}C site in only one strand, as indicated by the asterisk. Methylation with the other enzymes yields 5′-Pu^{m}C dinucleotides in both strands of the MTase target sequences with differences, however, in their relative position; the dinucleotides are directly opposed in the case of M.*Hae*III methylation, while they are slightly staggered in the case of M.*Hha*I methylation.

We first analysed the DNA binding capacity of the individual subunits of the McrBC restriction enzyme. Substrate DNA was used either non-methylated or methylated at its two M.*Hae*III sites and its unique M.*Msp*I site. All binding reactions were carried out in the presence of 1 mM GTP. Figure 4 shows that at low protein concentrations McrB binds to the modified DNA substrate yielding a defined fast migrating complex. With increasing McrB concentrations slow complexes appear, which form a 'diffuse' retardation band. This indicates the formation of high molecular weight complexes with a less defined stoichiometry of protein and DNA molecules at high McrB concentrations. As is evident from Figure 4, unmodified DNA also becomes bound by McrB at elevated concentrations. In contrast, however, to the slow migrating complexes formed with modified DNA at these McrB concentrations, only fast complexes forming defined retardation bands are observed.

No binding of the GST–McrC protein to either methylated or unmethylated DNA could be detected (Figure 4). Neither did GST–McrC modulate the DNA binding properties of McrB, as no effect on McrB was observed when GST–McrC was added to McrB at different ratios (data not shown).

To investigate whether GTP would affect DNA binding by McrB, increasing amounts of McrB protein were incubated with either the unmodified or the M.*Hae*III/M.*Msp*I-specifically modified DNA substrate in the presence or absence of 1 mM GTP. GTP enhanced the affinity of McrB for unmethylated, as well as for methylated, DNA (Figure 5). According to our estimations (which were based on a comparison of the amount of McrB protein required to achieve 50% binding of the substrate in the presence or absence of GTP respectively), the affinity of McrB for methylated DNA was increased ~2- to 3-fold by GTP. No quantitative estimation for binding to unmethylated DNA could be derived from our experiments, as within the McrB concentration range employed unmethylated DNA was not retarded in the absence of GTP. Whereas McrB apparently binds methylated DNA in a positively cooperative manner in the presence and absence of GTP (Figures 4 and 5A), a similar behaviour of McrB with non-methylated DNA in the presence of GTP, as suggested by Figure 5B, was only observed in some experiments. Further analyses will therefore be required to clarify the precise mode of interaction in this particular case. ATP neither enhanced nor inhibited binding of methylated DNA by McrB.

In a further series of experiments we investigated

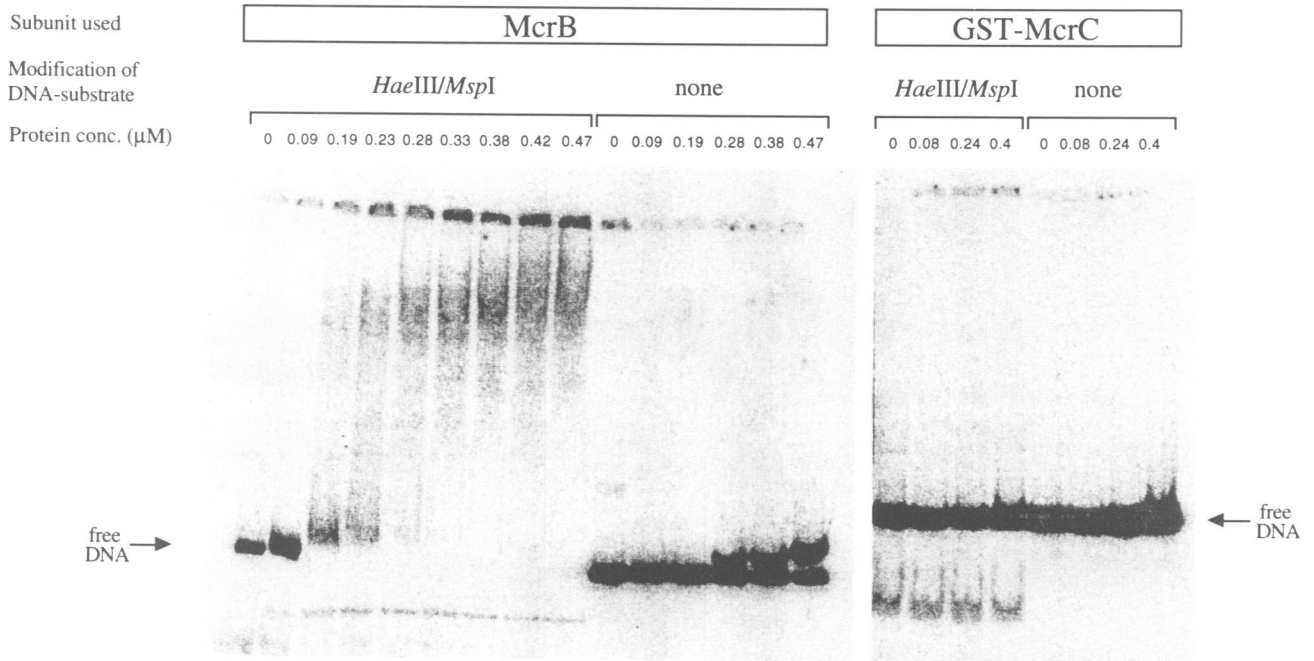

Fig. 4. Binding properties of the McrB and GST–McrC proteins with unmodified (none) and *Hae*III/*Msp*I-specifically modified DNA. The protein concentrations used are indicated. The DNA concentration in the reaction samples was ~0.1 nM. Arrows at the left and right side indicate the position of the unbound DNA fragments. The band below the free *Hae*III/*Msp*I-specifically modified DNA probe represents a labelled contaminant fragment of ~110 bp that could not be removed from the preparation prior to the labelling reaction.

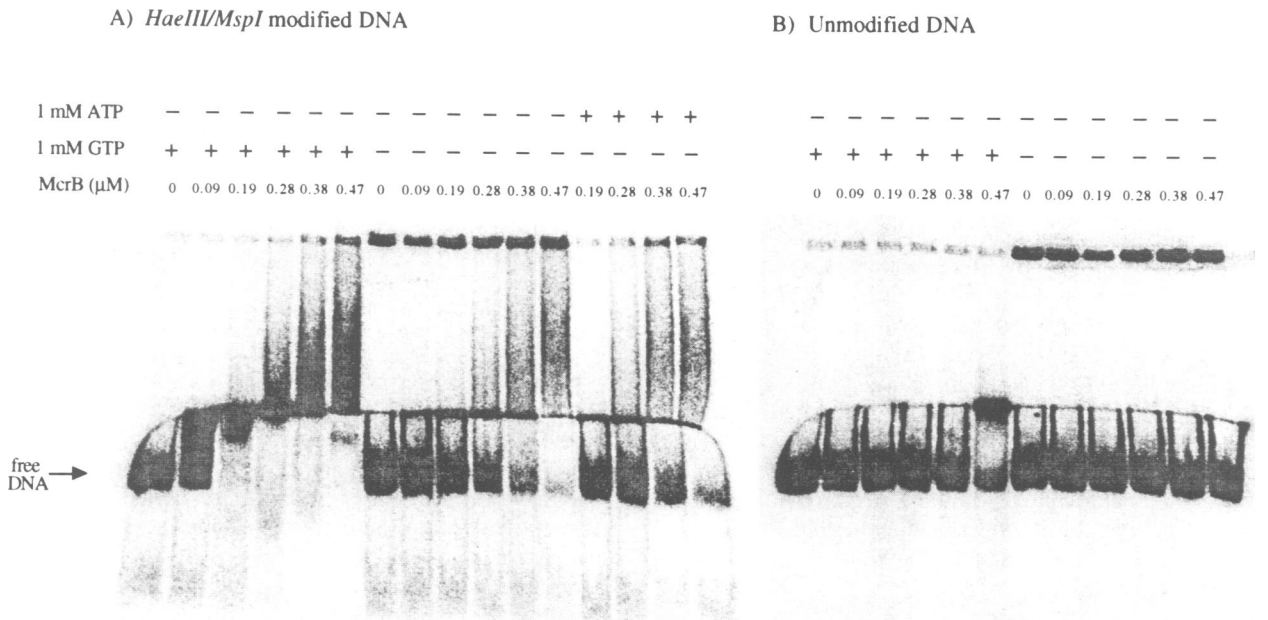

Fig. 5. Effect of GTP on the affinity of McrB for unmethylated and methylated DNA assayed by gel retardation. (**A**) Increasing concentrations of McrB protein were incubated with the unmethylated 241 bp fragment in the presence or absence of 1 mM GTP. The arrow indicates the position of the unbound DNA fragment. (**B**) Increasing concentrations of McrB protein were incubated with the *Hae*III/*Msp*I-specifically methylated fragment with or without 1 mM GTP. In the binding reactions assayed in lanes 13–16, GTP was replaced by 1 mM ATP. The sharp zone above the free DNA was only observed with a particular batch of the loading dye and can therefore be considered as a gel artifact.

whether and how differences in the number of 5′-Pu^{m}C dinucleotides within a given DNA fragment would affect binding by McrB. The 241 bp DNA substrate was used either in its unmethylated state or after selective modification by M.*Hha*I, M.*Hae*III or M.*Hae*III + M.*Msp*I (see Figure 3). Figure 6 shows that within the McrB concentration range used, neither the unmethylated nor the M.*Hha*I-specifically modified fragment became bound by McrB. The fragment with two methylated *Hae*III sequences became significantly retarded and the fragment additionally methylated at the *Msp*I site was bound even more efficiently. The latter result shows that a 5′-PuC site can be efficiently recognized by the McrB protein even when it is methylated in only one strand.

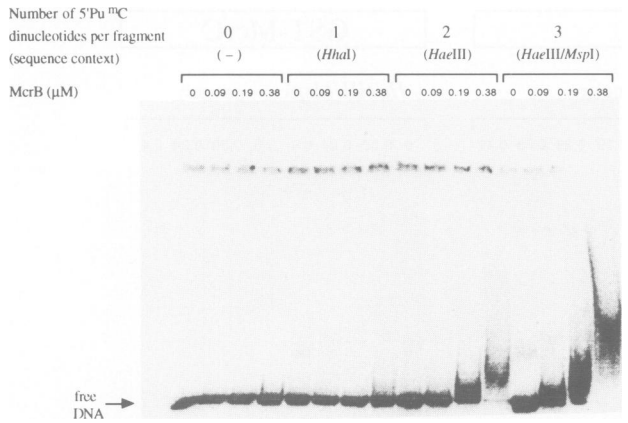

Fig. 6. Binding of McrB protein to DNA fragments with different numbers of 5′-Pu^{m}C sites. The protein concentrations in the binding reactions and the DNA fragment used (with the number of 5′-Pu^{m}C sites indicated) are shown above the lanes. The lanes labelled 0 contain only the binding substrate without protein. The arrow indicates the position of the unbound DNA fragment.

We wondered whether the inability of McrB to bind to M.*Hha*I-modified substrate was due to a general failure of the protein to recognize isolated 5′-Pu^{m}C sites or whether it depended on the particular sequence context found there. To address this point, we converted the central M.*Hae*III recognition sequence of the 241 bp fragment into a M.*Hha*I site by oligonucleotide-directed mutagenesis, generating a fragment with a single M.*Hae*III site and two *M.Hha*I sites (Figure 3B).

To allow a direct comparison, the McrB protein was incubated in parallel with the original and mutant 241 bp fragments which had been methylated with M.*Hha*I or M.*Hae*III (Figure 7A and B). As is evident from Figure 7, the original fragment methylated at its two *Hae*III sites is bound most strongly by McrB, followed by the mutant fragment methylated at its single *Hae*III site. The mutant fragment with two methylated *Hha*I sites is bound less efficiently and only very weak binding to the original fragment with only one methylated *Hha*I site can be observed. No binding to the totally unmodified fragment could be detected at the McrB concentrations employed. These results show that a single 5′-Pu^{m}C site can be sufficient to significantly stimulate McrB binding. It is also evident that the sequence context of 5′-Pu^{m}C sites modulates their signal character for McrB.

Discussion

The functional characterization of the McrBC nuclease described here involved purified GST–McrB/C fusion proteins and a preparation of McrB consisting of a roughly equimolar mixture of two McrB derivatives whose N-termini are slightly elongated relative to the N-terminus proposed by Dila *et al.* (1992). DNA restriction experiments revealed the qualitative identity of these proteins and their native counterparts: restriction required the combination of derivatives of both McrBC subunits, depended on DNA modifications eliciting restriction by the native proteins and implied availability of GTP as cofactor in the reaction. In addition, DNA cleavage with the purified proteins yielded the same fragment patterns,

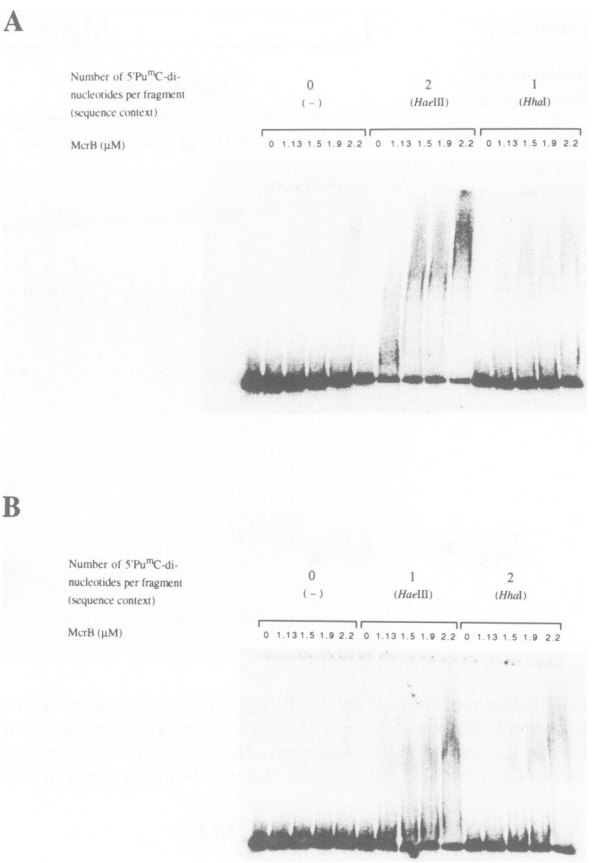

Fig. 7. Effect of the number and sequence context of 5′-Pu^{m}C sites on McrB binding. The binding substrate in the experiments shown in (**A**) is the 241 bp fragment shown in Figure 3A, containing two *Hae*III and one *Hha*I recognition sequences. In the experiment shown in (**B**) the binding substrates were differentially methylated derivatives of the mutated 241 bp fragment with one *Hae*III and two *Hha*I sites (Figure 3B). The numbers of symmetrically modified 5′-Pu^{m}C sites in methylated derivatives of the fragment and their immediate sequence contexts are indicated. The absolute McrB concentrations necessary to obtain the retardation effect were significantly higher compared with the previous experiments, due to a somewhat lower DNA binding activity of the protein preparation.

with subtle differences in the intensity of individual bands, as characteristically generated by the native proteins (Sutherland *et al.*, 1992; this communication). The qualitative equivalence of native proteins and fusion derivatives, as documented here, is not unusual. It has been demonstrated in several other cases, in particular in studies on prokaryotic endonucleases (Bhagwat *et al.*, 1990; Hennecke *et al.*, 1991) and a eukaryotic methylated-DNA binding protein (Lewis *et al.*, 1992). Insensitivity of proteins to small additions or truncations at their N-termini has also frequently been observed.

Our analyses lead to several conclusions concerning the general functional organization of the McrBC nuclease. Evidently, DNA binding by the enzyme is mediated mainly or exclusively by the McrB protein. This protein can, hence, be considered as the DNA binding subunit of the McrBC nuclease. McrC, which due to its overall basic character (basic residues are neither confined to nor concentrated within a particular region) has been suggested to be the DNA binding subunit (Dila *et al.*, 1990), neither revealed DNA binding potential nor detectably modulated

McrB binding to DNA. As McrC is an obligatory requirement for restriction, it appears that this protein component essentially contributes to the catalytic capacity of the McrBC nuclease and may actually represent its catalytic subunit.

The finding that GTP stimulates binding of McrB to DNA is the first experimental evidence provided for a GTP binding potential of McrB. Binding of GTP by McrB had been suggested before by Dila *et al.* (1990), who identified a tripartite motif in McrB resembling similarly organized motifs of various GTP binding proteins (Dever *et al.*, 1987). The rather weak stimulatory effect of GTP on DNA binding by McrB, however, does not explain the essential role of this cofactor in the DNA restriction reaction, which, as deduced from experiments with non-hydrolysable analogues, involves its hydrolysis (Sutherland *et al.*, 1992). In addition to stimulating DNA binding by McrB, GTP apparently plays another important role in the restriction reaction. In contrast to previous reports (Sutherland *et al.*, 1992), we could not demonstrate an inhibitory effect of ATP on McrBC restriction. Consequently, no attempts were made to assign ATP binding to any of the two McrBC subunits.

Several of the general properties of the McrBC restriction endonuclease described here and before (Sutherland *et al.*, 1992) are reminiscent of type I restriction endonucleases (DNA binding and restriction are conferred by different subunits; binding at two target sites is a prerequisite for restriction; DNA cleavage strictly depends on nucleotide cofactors). In view of the different nucleotide cofactors employed by these nucleases (ATP by type I restriction endonucleases, GTP by McrBC), however, it is questionable whether these formal similarities reflect any relatedness in the mechanistic details of DNA binding and DNA restriction. While ATP, which usually serves as a mere energy donor in biochemical reactions, provides the energy to translocate the intervening DNA between two sites bound by type I endonuclease molecules (Studier and Bandyopadhyay, 1988), GTP usually serves in the controlled transition of protein conformations from inactive to active states which occurs as a consequence of GTP hydrolysis (Bourne *et al.*, 1990, 1991).

Most of the properties of McrBC distinguish this nuclease from the only other biochemically well characterized, methylation-dependent endonuclease, R.*Dpn*I from *Streptococcus pneumoniae* (Lacks and Greenberg, 1975; Lacks, 1980). Besides its requirement for adenine *N*6-methylation at GATC sites, R.*Dpn*I exhibits features characteristic of type II restriction endonucleases.

The details of the DNA binding properties of McrB are of particular interest in two respects: how do they correlate with DNA restriction by the McrBC endonuclease and how do they compare with the DNA binding properties of eukaryotic proteins that bind to methylated DNA? The discrimination by the McrBC nuclease between non-modified (non-restrictable) and modified (restrictable) DNA substrates can be attributed to the DNA binding properties of McrB. The protein has a lower affinity for non-modified DNA substrates than for their modified counterparts. Furthermore, the complexes formed with unmodified and modified DNA substrates are strikingly different. With unmodified DNA low molecular weight complexes are formed, while the interaction with modified substrates yields high molecular weight complexes, presumably comprising several McrB molecules. It is conceivable that in addition to differences in the affinity, differences in the quality of complexes formed might also contribute to the discrimination between unmethylated and methylated DNA in McrBC restriction.

The finding that 5′-PuC sites methylated in only one, as well as those methylated in both, strands can be efficiently recognized by the McrB protein also parallels DNA restriction by the McrBC enzyme. As shown by Sutherland *et al.* (1992), hemimethylation at two appropriately spaced 5′-PuC sites is sufficient to elicit McrBC restriction. A discussion of analogies between DNA restriction by McrBC and DNA binding by McrB should also consider the observed substantial modulation of the signal character of 5′-Pu^{m}C sites caused by the sequence context. From the experiments described here we cannot discriminate between this modulation depending on general parameters of the sequence environment and/or subtle differences in the constellation of 5′-Pu^{m}C sites in opposite strands of DNA MTase target sequences. A major contribution of general parameters, however, appears likely, in view of the different intensities of individual bands in patterns generated by McrBC restriction of uniformly methylated DNA substrates (Sutherland *et al.*, 1992; see also Figure 2).

A striking difference to McrBC restriction, which depends on the presence of at least two 5′-Pu^{m}C sites, is the finding that a single 5′-Pu^{m}C site can provide an efficient signal for McrB binding. As far as is deducible from our experiments, there is no obvious difference in the quality of complexes formed with DNA substrates containing only a single or at least two 5′-Pu^{m}C sites. In both cases high molecular weight complexes are obtained. What prevents restriction of such substrate DNA by the McrBC enzyme? Apparently coordinated interaction between McrB(C?) molecules simultaneously bound at two appropriately spaced 5′-Pu^{m}C sites is of major importance for restriction. This interaction could, similarly to that described for type I restriction endonucleases, be brought about by looping out the intervening DNA sequence. Alternatively, taking into account the limited spacing allowed for 5′-Pu^{m}C sites and the capacity of McrB to form high molecular weight complexes, it might be achieved by bridging the intervening sequence with McrB(C?) molecules.

In a discussion of the potential evolutionary origin of methylation-associated gene inactivation in eukaryotes, Bestor (1990) proposed that the components of a possibly underlying system, C5-DNA MTases and proteins specifically binding to cytosine-methylated DNA, might both have evolved from prokaryotic ancestors. In fact, the C-terminal parts of the murine and human C5-MTases strongly resemble prokaryotic type II C5-MTases (Bestor *et al.*, 1988; Yen *et al.*, 1992), indicating a common evolutionary origin of these enzymes.

McrB, which binds to DNA containing modified cytosine residues, might share a common ancestor with eukaryotic proteins binding to methylated DNA. Its mode of DNA binding is basically different from that of the ubiquitous vertebrate protein MDBP-1 (Supakar *et al.*, 1988; Ehrlich and Ehrlich, 1993), which binds to methylated DNA in a sequence-specific manner, i.e. only when

5′-^{m}CG dinucleotides occur within a larger defined base context. DNA binding by McrB resembles in several respects, however, that by the vertebrate proteins MeCP1, MeCP2 and MDBP-2 (Meehan *et al.*, 1989, 1992; Jost and Hofsteenge, 1992; Lewis *et al.*, 1992; Nan *et al.*, 1993) and the plant protein DBP-m (Zhang *et al.*, 1989; Ehrlich and Ehrlich, 1993). These proteins bind to DNA in response to methylated di- or trinucleotides (5′-Pu^{m}C in the case of McrB; 5′-^{m}CG in the case of MeCP1, MeCP2 and MDBP-2; 5′-^{m}CG, 5′-^{m}CNG and 5′-^{m}CTA in the case of DBP-m) in an otherwise virtually sequence-independent manner. For McrB, MeCP2, MDBP-2 and DBP-m single sites of the type specified above are sufficient to allow binding (at least 12 such sites are required to permit efficient binding of MeCP1 to oligomers; Meehan *et al.*, 1989). In addition, McrB and DBP-m share the capacity to efficiently bind to their hemimethylated di- or trinucleotide targets (Zhang *et al.*, 1989; this communication). In this respect they differ from the other proteins discussed, which require symmetrically methylated 5′-CG sites for efficient binding (Meehan *et al.*, 1989; Jost and Hofsteenge, 1992; Lewis *et al.*, 1992).

In view of only limited information on the primary sequences of the proteins dicussed and in the absence of knowledge of their tertiary structures, it is hard to judge whether the general similarities discussed reflect an evolutionary relationship. Comparing the amino acid sequence of McrB with that of MeCP2 (Lewis *et al.*, 1992), the only eukaryotic methylated DNA binding protein whose primary sequence is known, revealed no significant similarities. The lack of similarity at the primary sequence level, however, does not necessarily argue against an evolutionary relatedness. The type II restriction endonucleases R.*Bam*HI and R.*Eco*RI, for example, which share no identity in their primary sequences, are strikingly similar in their three-dimensional architecture. This led to the proposal of a model implying their evolution from a common ancestor (Newman *et al.*, 1994). Further sequence and structure information has to be awaited to decide whether the eukaryotic proteins binding to methylated DNA have prokaryotic roots, such as McrB.

Materials and methods

Bacterial strains, phages and media

The methylation-tolerant *E.coli* strain TC410 [*mcrA*$^{-}$, Δ(*mrr-hsdRMS-mcrBC*)201, *minA*, *minB*, *rpsL*, sup^{+}; Noyer-Weidner *et al.*, 1986; Krüger *et al.*, 1992] served as a general cloning host. JM109 [*recA*1, *supE*44, *endA*1, *hsdR*17, *gyrA*96, *relA*1, *thi*, Δ(*lac-proAB*), (F′, *traD*36, *proAB*, *lacI*^{q}, ZΔM15; Yanisch-Perron *et al.*, 1985] was used in cloning experiments employing M13mp18/mp19 vectors. Cells of strain ER1647 [F^{-}, λ^{-}, *trp31*, *his1*, *fhu1*, Δ(*lacZ*)*r1*, *rpsL104*, *supE44*, Δ(*mcrC-mrr*)*102*::Tn*10*, *mcrA1272*::Tn*10*, *recD1014*, *xyl7*, *mtl2*, *metB1*; Sutherland *et al.*, 1992] were transformed with *mcrB/C*-expressing plasmids and used for crude cell extract preparation. All *E.coli* strains were grown in LB medium supplemented with 50 μg/ml ampicillin, if necessary.

Enzymes and radiochemicals

Restriction endonucleases, the Klenow fragment of DNA polymerase I, calf intestine phosphatase and protease Factor Xa were from Boehringer (Mannheim, Germany). T4 DNA ligase and T4 polynucleotide kinase were obtained from BRL (Neu Isenburg, Germany). Protease thrombin was from Sigma (München, Germany) and DNA MTases M.*Hha*I, M.*Msp*I and M.*Hae*III were purchased from New England Biolabs (Beverly, MA). All enzymes were used according to the manufacturers recommendations. [α-^{35}S]dATP and [γ-^{32}P]ATP were obtained from Amersham-Buchler (Braunschweig, Germany).

Molecular cloning procedures

Preparation of plasmid DNA, enzymatic digestion of DNA, isolation of DNA fragments, blunting of fragment ends, DNA ligation and transformation of competent *E.coli* cells were performed as described by Sambrook *et al.* (1989).

Construction of plasmids

To create a fusion of the *gst* gene with the *mcrB* gene, a 1.5 kb *Esp*I–*Nsi*I subfragment was isolated from the 2.7 kb *Hpa*I–*Stu*I fragment covering the whole *mcrBC* region of *E.coli* (Krüger *et al.*, 1992). After blunt-ending, this fragment was inserted into the *Eco*RI site of pGEX-2T, which had previously been filled in with Klenow polymerase, to yield pBN210. For construction of the *gst–mcrC* fusion, a 1.2 kb *Ava*II–*Stu*I subfragment was blunt-ended and inserted into the Klenow polymerase-treated *Eco*RI site of pGEX2T, creating plasmid pBN212. The *gst–mcrC* fusion containing a factor Xa cleavage site was constructed by inserting the blunt-ended *Ava*II–*Stu*I fragment into the *Sma*I site of pGEX-3X (pBN213). The junction between *gst* and *mcrB* in plasmid pBN210 was corrected by oligonucleotide-directed mutagenesis: a 211 bp *Bam*HI–*Hin*dIII fragment covering the fusion point was cloned into M13mp19 and single-stranded DNA of the resulting derivative was prepared. Five picomoles of phosphorylated oligonucleotide (26mer) was annealed to 1 μg ssDNA and the second strand was synthesized in the presence of all four dNTPs and 3 U Klenow polymerase and T4 DNA ligase. The DNA was transformed into competent JM109 cells and single plaques were screened for the presence of a newly introduced *Ssp*I site. Cloning of the corrected 211 bp fragment into *Bam*HI/*Hin*dIII-digested pBN210 yielded plasmid pBN211.

Crude cell extract preparation and DNA restriction assay

TC410 cells harbouring the appropriate plasmid (pBN210/pBN212) were cultivated in a 20 ml volume to an OD_{560} of 0.8–1.0, protein expression was induced with 0.5 mM IPTG and the cultures were grown for a further 3 h. The cells were pelleted by centrifugation (Sorvall HB4 rotor, 10 min at 6000 r.p.m.), the pellet was washed with 10 ml HEPES buffer [20 mM HEPES, pH 7.6, 50 mM KCl, 1 mM EDTA, 1 mM dithiothreitol, 10% (v/v) glycerol] and resuspended in 2 ml of the same buffer. Lysis of the cells was achieved by sonication on ice (volume 1 ml, 8×10 s with a Ultrasonic Disintegrator from MSE, London, UK) and 0.1 ml aliquots of the extracts were stored at –18°C.

In vitro McrB/C restriction of specifically methylated substrate DNA was performed at 37°C. The reaction mixtures contained 0.5 or 1 μg substrate DNA, 5 mM $MgCl_2$, 1 mM GTP, 50 mM KCl and 20 mM HEPES–KOH, pH 7.6, in a total volume of 20 μl. Reactions were stopped on ice, mixed with 4 μl loading dye and electrophoresed through a 1% agarose gel at 10 V/cm in TAE buffer. The gels were stained in TAE buffer containing 2 μg/ml ethidium bromide. McrB- or McrC-containing crude extracts (2 μl/sample) were used and reactions with purified proteins contained 0.1 μg McrB and 0.3 μg GST–McrC respectively.

Protein purification

For analytical detection of protein overexpression, liquid cultures of transformed cells were grown in LB medium to an OD_{560} of 1.0 and samples 'before induction' were taken. Protein expression was induced with 0.5 mM IPTG for 2 h, after which the cells were harvested and 10-fold concentrated in 20 mM HEPES, pH 7.6, 1 mM EDTA, 50 mM KCl, 10% (v/v) glycerol. An aliquot was made 100 mM in NaCl and 2.5 mM in $CaCl_2$ and incubated for 30 min at room temperature with 0.5 μg thrombin (Sigma, München). An amount equal to 100 μl of culture was mixed with loading buffer, heat denatured and electrophoresed through a 15% polyacrylamide gel, which was stained with Coomassie Blue (Serva, Heidelberg, Germany).

During preparative protein purification all steps were carried out at 4°C or on ice, unless otherwise stated. For McrB, two 1 l cultures of cells containing pBN211 were grown in LB medium (supplemented with 50 μg/ml ampicillin) to an OD_{560} of 1.0 and induced with 0.5 mM IPTG for 3 h with shaking at 30°C to prevent the formation of inclusion bodies. Cells were harvested by centrifugation (Sorvall GSA rotor, 10 min at 6000 r.p.m.), washed once with HEPES buffer and resuspended in 20 ml (1/100 culture volume) of the same buffer. The cells were disrupted in a French press cell at 16 000 p.s.i. and the cell debris was pelleted in a Sorvall SS34 rotor for 30 min at 15 000 r.p.m.. The supernatant was diluted with half the volume of HEPES buffer and applied onto a pre-equilibrated 2 ml column of glutathione–Sepharose 4B (Pharmacia, Uppsala, Sweden). After washing the column with 10 column volumes HEPES buffer, 500 mM KCl, the GST–McrB fusion

protein was eluted with 10 mM glutathione in HEPES buffer (pH adjusted to 7.5). Fusion protein-containing fractions were pooled and made 100 mM in NaCl and 2.5 mM in $CaCl_2$. Proteolysis of the fusion protein with thrombin (30 min at room temperature, ratio by weight 1:200) was followed by 3-fold dilution of the sample with HEPES buffer without KCl and application onto a pre-equilibrated 2 ml heparin–Sepharose column (Pharmacia, Uppsala, Sweden). After washing the column with 20 vol. HEPES buffer, 150 mM KCl, the McrB protein was eluted with HEPES buffer, 300 mM KCl. McrB-containing fractions were pooled, dialysed against HEPES buffer, concentrated with Amicon concentrators and stored at −18°C. For GST–McrC the purification scheme was identical to that described for the McrB protein until the glutathione elution step. The GST–McrC-containing eluate was pooled, dialysed against HEPES buffer, concentrated and stored at −18°C. Protein concentrations were determined with a protein assay kit from Bio-Rad (München, Germany), based on the Bradford method.

N-Terminal protein sequencing

After cleavage of the GST–McrB fusion protein with thrombin, a sample with ~80 mg McrB protein was electrophoresed through a 15% polyacrylamide gel for 6 h for complete separation of the two McrB product bands. The proteins were electro-blotted onto a PVDF membrane (Millipore, Eschborn, Germany) at 180 mA and 4°C. Protein bands were visualized on the membrane with Ponceau S solution (Sigma, München, Germany), cut out and the membrane pieces destained with phosphate-buffered saline. The N-terminal degradations were performed on a pulsed liquid-phase sequencer (Applied Biosystems model 447A) and the phenylthiohydantoin derivatives of the amino acids were separated on a model 120A analyser from Applied Biosystems.

Gel retardation

DNA fragments were prepared by amplification of the appropriate region of the plasmid pMS119EH by the polymerase chain reaction (PCR) using oligonucleotide primers corresponding to nucleotides 156–180 of one DNA strand of the plasmid and nucleotides 386–411 of the complementary strand respectively. The PCR reactions were phenol extracted, ethanol precipitated and sequentially digested with *Ava*II and *Acy*I. Fragments were purified on a 5% polyacrylamide gel, dephosphorylated with calf intestine phosphatase and methylated *in vitro* with either M.*Hae*III, M.*Msp*I or M.*Hha*I. Approximately 10–200 ng of the methylated fragments were used in labelling reactions with T4 polynucleotide kinase and [γ-^{32}P]ATP (5000 Ci/mmol), followed by removal of unincorporated nucleotides by gel filtration on G-50 Quick-Spin columns (Boehringer, Mannheim, Germany). The volume activities were measured by Cherenkov counting in a Beckman LS 7800 liquid scintillation counter. Completeness of the *in vitro* methylation reactions was verified by the absence of digestion products after incubation of the labelled DNA fragments with an excess of the appropriate restriction endonuclease followed by gel electrophoresis and autoradiography. The standard binding reactions contained 20 mM HEPES–KOH, pH 7.6, 1 mM GTP, 50 mM KCl, 1000–2000 c.p.m. (Cherenkov counts) radiolabelled DNA fragment (corresponding to a final DNA concentration of ~0.1 nM) and the indicated amounts of McrB/GST–McrC protein in a total volume of 20 µl. Incubation was for 15 min at room temperature, after which the reactions were mixed with 4 µl loading dye (type I; Sambrook *et al.*, 1989), immediately loaded onto 5% polyacrylamide gels and electrophoresed for 10 min at 20 V/cm and 2 h at 10 V/cm in TBE buffer. After electrophoresis, gels were transferred onto sheets of Whatman 3MM paper, dried under vacuum and exposed to a PhosphorImager screen at room temperature. The screen was scanned with a PhosphorImager from Molecular Dynamics (Sunnyvale, CA) and the images were created with the ImageQuant software.

Acknowledgements

We would like to thank Hans-Joachim Fritz, Walter Schaffner and Jörn Walter for critical reading of the manuscript and helpful comments. We gratefully acknowledge discussions with Alfred Pingoud. We are furthermore indebted to Thomas A.Trautner for his support and many stimulating discussions. M.N.-W. was supported by grant Tr25/10-3 of the Deutsche Forschungsgemeinschaft.

References

Adhin,M.R. and van Duin,J. (1990) *J. Mol. Biol.*, **213**, 811–818.
Bestor,T.H. (1990) *Phil. Trans. R. Soc. B*, **326**, 179–187.
Bestor,T., Laudano,A., Mattaliano,R. and Ingram,V. (1988) *J. Mol. Biol.*, **203**, 971–983.
Bhagwat,A.S., Johnson,B., Weule,K. and Roberts,R.J. (1990) *J. Biol. Chem.*, **265**, 767–773.
Bourne,H.R., Sanders,D.A. and McCormick,F. (1990) *Nature*, **348**, 125–132.
Bourne,H.R., Sanders,D.A. and McCormick,F. (1991) *Nature*, **349**, 117–127.
Chang,J.-Y. (1985) *Eur. J. Biochem.*, **151**, 217–224.
Dever,T.E., Glynias,M.J. and Merrick,W.C. (1987) *Proc. Natl Acad. Sci. USA*, **84**, 1814–1818.
Dila,D., Sutherland,E., Moran,L., Slatko,B. and Raleigh,E.A. (1990) *J. Bacteriol.*, **172**, 4888–4900.
Ehrlich,M. and Ehrlich,K.C. (1993) In Jost,J.P. and Saluz,H.P. (eds), *DNA Methylation: Molecular Biology and Biological Significance*. Birkhäuser Verlag, Basel, Switzerland, pp. 145–168.
Fürste,J.P., Pansegrau,W., Frank,R., Blöcker,H., Scholz,P., Bagdasarian,M. and Lanka,E. (1986) *Gene*, **48**, 119–131.
Hennecke,F., Kolmar,H., Bründl,K. and Fritz,H.-J. (1991) *Nature*, **353**, 776–778.
Jost,J.-P. and Hofsteenge,J. (1992) *Proc. Natl Acad. Sci. USA*, **89**, 9499–9503.
Krüger,T., Grund,C., Wild,C. and Noyer-Weidner,M. (1992) *Gene*, **114**, 1–12.
Lacks,S.A. (1980) *Methods Enzymol.*, **65**, 138–146.
Lacks,S. and Greenberg,B. (1975) *J. Biol. Chem*, **250**, 4060–4066.
Lewis,J.D., Meehan,R.R., Henzel,W.J., Maurer-Fogy,I., Jeppesen,P., Klein,F. and Bird,A. (1992) *Cell*, **69**, 905–914.
Meehan,R.R., Lewis,J.D., McKay,S., Kleiner,E.L. and Bird,A.P. (1989) *Cell*, **58**, 499–507.
Meehan,R.R., Lewis,J.D. and Bird,A.P. (1992) *Nucleic Acids Res.*, **20**, 5085–5092.
Nagai,K. and Thøgersen,H.C. (1984) *Nature*, **309**, 810–812.
Nan,X., Meehan,R.R. and Bird,A.P. (1993) *Nucleic Acids Res.*, **21**, 4886–4892.
Newman,M., Strzelecka,T., Dorner,L.F., Schildkraut,I. and Aggarwal, A.K. (1994) *Nature*, **368**, 660–664.
Noyer-Weidner,M. and Trautner,T.A. (1993) In Jost,J.P. and Saluz,H.P. (eds), *DNA Methylation: Molecular Biology and Biological Significance*. Birkhäuser Verlag, Basel, Switzerland, pp. 39–108.
Noyer-Weidner,M., Diaz,R. and Reiners,L. (1986) *Mol. Gen. Genet.*, **205**, 469–475.
Raleigh,E.A. (1992) *Mol. Microbiol.*, **6**, 1079–1086.
Raleigh,E.A., Trimarchi,R. and Revel,H. (1989) *Genetics*, **122**, 279–296.
Ross,T.K., Achberger,E.C. and Braymer,H.D. (1989) *J. Bacteriol.*, **171**, 1974–1981.
Sambrook,F., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Schümperli,D., McKenney,K., Sobieski,D.A. and Rosenberg,M. (1982) *Cell*, **30**, 865–871.
Smith,D.B. and Johnson,K.S. (1988) *Gene*, **67**, 31–40.
Studier,F.W. and Bandyopadhyay,P.K. (1988) *Proc. Natl Acad. Sci. USA*, **85**, 4677–4681.
Supakar,P.C., Weist,D., Zhang,D., Inamdar,X.-Y., Khan,R., Ehrlich,K.C. and Ehrlich,M. (1988) *Nucleic Acids Res.*, **16**, 8029–8044.
Sutherland,E., Coe,L. and Raleigh,E.A. (1992) *J. Mol. Biol.*, **225**, 327–348.
Walter,J. (1990) PhD Thesis, Free University of Berlin.
Yanisch-Perron,C., Vieira,J. and Messing,J. (1985) *Gene*, **33**, 103–119.
Yen,R.-W.C. *et al.* (1992) *Nucleic Acids Res.*, **20**, 2287–2291.
Zhang,D., Ehrlich,K.C., Supakar,P.C. and Ehrlich,M. (1989) *Mol. Cell. Biol.*, **9**, 1351–1356.

Received on January 11, 1995; revised on March 13, 1995